

RAW SEQUENCE LISTING

**The Biotechnology Systems Branch of the Scientific and Technical
Information Center (STIC) no errors detected.**

Application Serial Number: 10/775, 678A
Source: IFW/6
Date Processed by STIC: 03/14/2007

ENTERED



IFW16

RAW SEQUENCE LISTING

DATE: 03/14/2007

PATENT APPLICATION: US/10/775,678A

TIME: 10:46:31

Input Set : A:\10278-048001.txt

Output Set: N:\CRF4\03142007\J775678A.raw

3 <110> APPLICANT: Transkaryotic Therapies, Inc.
 4 von Figura, Kurt
 5 Schmidt, Bernhard
 6 Dierks, Thomas
 7 Heartlein, Michael W.
 8 Cosma, Maria P.
 9 Ballabio, Andrea
 11 <120> TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MULTIPLE SULFATASE DEFICIENCY AND
 12 OTHER SULFATASE DEFICIENCIES
 14 <130> FILE REFERENCE: 10278-048001
 C--> 16 <140> CURRENT APPLICATION NUMBER: US/10/775,678A
 17 <141> CURRENT FILING DATE: 2004-02-10
 19 <150> PRIOR APPLICATION NUMBER: US 60/447,747
 20 <151> PRIOR FILING DATE: 2003-02-11
 22 <160> NUMBER OF SEQ ID NOS: 96
 24 <170> SOFTWARE: PatentIn version 3.2
 26 <210> SEQ ID NO: 1
 27 <211> LENGTH: 1180
 28 <212> TYPE: DNA
 29 <213> ORGANISM: Homo sapiens
 32 <220> FEATURE:
 33 <221> NAME/KEY: CDS
 34 <222> LOCATION: (20)..(1141)
 36 <400> SEQUENCE: 1
 37 acatggcccg cgggacaac atg gct gcg ccc gca cta ggg ctg gtg tgt gga 52
 38 Met Ala Ala Pro Ala Leu Gly Leu Val Cys Gly
 39 1 5 10
 41 cgt tgc cct gag ctg ggt ctc gtc ctc ttg ctg ctg ctg ctc tcg ctg 100
 42 Arg Cys Pro Glu Leu Gly Leu Val Leu Leu Leu Leu Leu Ser Leu
 43 15 20 25
 45 ctg tgt gga gcg gca ggg agc cag gag gcc ggg acc ggt gcg ggc gcg 148
 46 Leu Cys Gly Ala Ala Gly Ser Gln Glu Ala Gly Thr Gly Ala Gly Ala
 47 30 35 40
 49 ggg tcc ctt gcg ggt tct tgc ggc tgc ggc acg ccc cag cgg cct ggc 196
 50 Gly Ser Leu Ala Gly Ser Cys Gly Cys Gly Thr Pro Gln Arg Pro Gly
 51 45 50 55
 53 gcc cat ggc agt tcg gca gcc gct cac cga tac tcg cgg gag gct aac 244
 54 Ala His Gly Ser Ser Ala Ala Ala His Arg Tyr Ser Arg Glu Ala Asn
 55 60 65 70 75
 57 gct ccg ggc ccc gta ccc gga gag cgg caa ctc gcg cac tca aag atg 292
 58 Ala Pro Gly Pro Val Pro Gly Glu Arg Gln Leu Ala His Ser Lys Met
 59 80 85 90
 61 gtc ccc atc cct gct gga gta ttt aca atg ggc aca gat gat cct cag 340

(pg-6)

RAW SEQUENCE LISTING

DATE: 03/14/2007

PATENT APPLICATION: US/10/775,678A

TIME: 10:46:31

Input Set : A:\10278-048001.txt

Output Set: N:\CRF4\03142007\J775678A.raw

62	Val	Pro	Ile	Pro	Ala	Gly	Val	Phe	Thr	Met	Gly	Thr	Asp	Asp	Pro	Gln	
63				95					100					105			
65	ata	aag	cag	gat	ggg	gaa	gca	cct	gcg	agg	aga	gtt	act	att	gat	gcc	388
66	Ile	Lys	Gln	Asp	Gly	Glu	Ala	Pro	Ala	Arg	Arg	Val	Thr	Ile	Asp	Ala	
67			110					115					120				
69	ttt	tac	atg	gat	gcc	tat	gaa	gtc	agt	aat	act	gaa	ttt	gag	aag	ttt	436
70	Phe	Tyr	Met	Asp	Ala	Tyr	Glu	Val	Ser	Asn	Thr	Glu	Phe	Glu	Lys	Phe	
71		125					130					135					
73	gtg	aac	tca	act	ggc	tat	ttg	aca	gag	gct	gag	aag	ttt	ggc	gac	tcc	484
74	Val	Asn	Ser	Thr	Gly	Tyr	Leu	Thr	Glu	Ala	Glu	Lys	Phe	Gly	Asp	Ser	
75	140				145				150				155				
77	ttt	gtc	ttt	gaa	ggc	atg	ttg	agt	gag	caa	gtg	aag	acc	aat	att	caa	532
78	Phe	Val	Phe	Glu	Gly	Met	Leu	Ser	Glu	Gln	Val	Lys	Thr	Asn	Ile	Gln	
79			160					165					170				
81	cag	gca	gtt	gca	gct	gct	ccc	tgg	tgg	tta	cct	gtg	aaa	ggc	gct	aac	580
82	Gln	Ala	Val	Ala	Ala	Ala	Pro	Trp	Trp	Leu	Pro	Val	Lys	Gly	Ala	Asn	
83			175					180				185					
85	tgg	aga	cac	cca	gaa	ggg	cct	gac	tct	act	att	ctg	cac	agg	ccg	gat	628
86	Trp	Arg	His	Pro	Glu	Gly	Pro	Asp	Ser	Thr	Ile	Leu	His	Arg	Pro	Asp	
87		190					195					200					
89	cat	cca	gtt	ctc	cat	gtg	tcc	tgg	aat	gat	gcg	gtt	gcc	tac	tgc	act	676
90	His	Pro	Val	Leu	His	Val	Ser	Trp	Asn	Asp	Ala	Val	Ala	Tyr	Cys	Thr	
91		205				210					215						
93	tgg	gca	ggg	aag	cgg	ctg	ccc	acg	gaa	gct	gag	tgg	gaa	tac	agc	tgt	724
94	Trp	Ala	Gly	Lys	Arg	Leu	Pro	Thr	Glu	Ala	Glu	Trp	Glu	Tyr	Ser	Cys	
95	220				225				230				235				
97	cga	gga	ggc	ctg	cat	aat	aga	ctt	ttc	ccc	tgg	ggc	aac	aaa	ctg	cag	772
98	Arg	Gly	Gly	Leu	His	Asn	Arg	Leu	Phe	Pro	Trp	Gly	Asn	Lys	Leu	Gln	
99			240					245				250					
101	ccc	aaa	ggc	cag	cat	tat	gcc	aac	att	tgg	cag	ggc	gag	ttt	ccg	gtg	820
102	Pro	Lys	Gly	Gln	His	Tyr	Ala	Asn	Ile	Trp	Gln	Gly	Glu	Phe	Pro	Val	
103			255					260				265					
105	acc	aac	act	ggt	gag	gat	ggc	ttc	caa	gga	act	gcg	cct	gtt	gat	gcc	868
106	Thr	Asn	Thr	Gly	Glu	Asp	Gly	Phe	Gln	Gly	Thr	Ala	Pro	Val	Asp	Ala	
107		270					275					280					
109	ttc	cct	ccc	aat	ggt	tat	ggc	tta	tac	aac	ata	gtg	ggg	aac	gca	tgg	916
110	Phe	Pro	Pro	Asn	Gly	Tyr	Gly	Leu	Tyr	Asn	Ile	Val	Gly	Asn	Ala	Trp	
111		285				290					295						
113	gaa	tgg	act	tca	gac	tgg	tgg	act	gtt	cat	cat	tct	gtt	gaa	gaa	acg	964
114	Glu	Trp	Thr	Ser	Asp	Trp	Trp	Thr	Val	His	His	Ser	Val	Glu	Glu	Thr	
115	300				305				310			315					
117	ctt	aac	cca	aaa	ggt	ccc	cct	tct	ggg	aaa	gac	cga	gtg	aag	aaa	ggt	1012
118	Leu	Asn	Pro	Lys	Gly	Pro	Pro	Ser	Gly	Lys	Asp	Arg	Val	Lys	Lys	Gly	
119			320					325				330					
121	gga	tcc	tac	atg	tgc	cat	agg	tct	tat	tgt	tac	agg	tat	cgc	tgt	gct	1060
122	Gly	Ser	Tyr	Met	Cys	His	Arg	Ser	Tyr	Cys	Tyr	Arg	Tyr	Arg	Cys	Ala	
123			335					340				345					
125	gct	cgg	agc	cag	aac	aca	cct	gat	agc	tct	gct	tcg	aat	ctg	gga	ttc	1108
126	Ala	Arg	Ser	Gln	Asn	Thr	Pro	Asp	Ser	Ser	Ala	Ser	Asn	Leu	Gly	Phe	

RAW SEQUENCE LISTING

DATE: 03/14/2007

PATENT APPLICATION: US/10/775,678A

TIME: 10:46:31

Input Set : A:\10278-048001.txt

Output Set: N:\CRF4\03142007\J775678A.raw

```

127          350          355          360
129 cgc tgt gca gcc gac cgc ctg ccc acc atg gac tgacaaccaa gggtagtctt 1161
130 Arg Cys Ala Ala Asp Arg Leu Pro Thr Met Asp
131          365          370
133 cccagtgcca aggagcagt 1180
136 <210> SEQ ID NO: 2
137 <211> LENGTH: 374
138 <212> TYPE: PRT
139 <213> ORGANISM: Homo sapiens
141 <400> SEQUENCE: 2
143 Met Ala Ala Pro Ala Leu Gly Leu Val Cys Gly Arg Cys Pro Glu Leu
144 1          5          10          15
147 Gly Leu Val Leu Leu Leu Leu Leu Ser Leu Leu Cys Gly Ala Ala
148          20          25          30
151 Gly Ser Gln Glu Ala Gly Thr Gly Ala Gly Ala Gly Ser Leu Ala Gly
152          35          40          45
155 Ser Cys Gly Cys Gly Thr Pro Gln Arg Pro Gly Ala His Gly Ser Ser
156          50          55          60
159 Ala Ala Ala His Arg Tyr Ser Arg Glu Ala Asn Ala Pro Gly Pro Val
160 65          70          75          80
163 Pro Gly Glu Arg Gln Leu Ala His Ser Lys Met Val Pro Ile Pro Ala
164          85          90          95
167 Gly Val Phe Thr Met Gly Thr Asp Asp Pro Gln Ile Lys Gln Asp Gly
168          100          105          110
171 Glu Ala Pro Ala Arg Arg Val Thr Ile Asp Ala Phe Tyr Met Asp Ala
172          115          120          125
175 Tyr Glu Val Ser Asn Thr Glu Phe Glu Lys Phe Val Asn Ser Thr Gly
176          130          135          140
179 Tyr Leu Thr Glu Ala Glu Lys Phe Gly Asp Ser Phe Val Phe Glu Gly
180 145          150          155          160
183 Met Leu Ser Glu Gln Val Lys Thr Asn Ile Gln Gln Ala Val Ala Ala
184          165          170          175
187 Ala Pro Trp Trp Leu Pro Val Lys Gly Ala Asn Trp Arg His Pro Glu
188          180          185          190
191 Gly Pro Asp Ser Thr Ile Leu His Arg Pro Asp His Pro Val Leu His
192          195          200          205
195 Val Ser Trp Asn Asp Ala Val Ala Tyr Cys Thr Trp Ala Gly Lys Arg
196          210          215          220
199 Leu Pro Thr Glu Ala Glu Trp Glu Tyr Ser Cys Arg Gly Gly Leu His
200 225          230          235          240
203 Asn Arg Leu Phe Pro Trp Gly Asn Lys Leu Gln Pro Lys Gly Gln His
204          245          250          255
207 Tyr Ala Asn Ile Trp Gln Gly Glu Phe Pro Val Thr Asn Thr Gly Glu
208          260          265          270
211 Asp Gly Phe Gln Gly Thr Ala Pro Val Asp Ala Phe Pro Pro Asn Gly
212          275          280          285
215 Tyr Gly Leu Tyr Asn Ile Val Gly Asn Ala Trp Glu Trp Thr Ser Asp
216          290          295          300
219 Trp Trp Thr Val His His Ser Val Glu Glu Thr Leu Asn Pro Lys Gly

```

RAW SEQUENCE LISTING

DATE: 03/14/2007

PATENT APPLICATION: US/10/775,678A

TIME: 10:46:31

Input Set : A:\10278-048001.txt

Output Set: N:\CRF4\03142007\J775678A.raw

```

220 305          310          315          320
223 Pro Pro Ser Gly Lys Asp Arg Val Lys Lys Gly Gly Ser Tyr Met Cys
224          325          330          335
227 His Arg Ser Tyr Cys Tyr Arg Tyr Arg Cys Ala Ala Arg Ser Gln Asn
228          340          345          350
231 Thr Pro Asp Ser Ser Ala Ser Asn Leu Gly Phe Arg Cys Ala Ala Asp
232          355          360          365
235 Arg Leu Pro Thr Met Asp
236          370
239 <210> SEQ ID NO: 3
240 <211> LENGTH: 1122
241 <212> TYPE: DNA
242 <213> ORGANISM: Homo sapiens
244 <400> SEQUENCE: 3
245 atggctgctgc cgcactagg gctggtgtgt ggacgttgcc ctgagctggg tctcgctctc 60
247 ttgctgctgc tgctctcgct gctgtgtgga gcggcagggg gccaggaggc cgggaccggt 120
249 gcgggcgcgg ggctcccttgc gggttcttgc ggctgcggca cgcgccagcg gcctggcgcc 180
251 catggcagtt cggcagccgc tcaccgatac tcgcgggagg ctaacgctcc gggccccgta 240
253 cccggagagc ggcaactcgc gcaactcaaag atgggtcccca tccctgctgg agtatattaca 300
255 atgggcacag atgactctca gataaagcag gatggggaag cacctgcgag gagagttact 360
257 attgatgcct ttacatgga tgccatgaa gtcagtaata ctgaatttga gaagtttgtg 420
259 aactcaactg gctatttgac agaggctgag aagtttgccg actcctttgt ctttgaaggc 480
261 atgttgagtg agcaagtga gaccaatatt caacaggcag ttgcagctgc tccctggtgg 540
263 ttacctgtga aaggcgctaa ctggagacac ccagaagggc ctgactctac tattctgcac 600
265 aggccggatc atccagttct ccatgtgtcc tgggaatgat cggttgccca ctgcacttgg 660
267 gcagggaagc ggctgcccac ggaagctgag tgggaataca gctgtcgagg aggctgcat 720
269 aatagacttt tcccctgggg caacaaactg cagcccaaag gccagcatta tgccaacatt 780
271 tggcagggcg agtttccggt gaccaacact ggtgaggatg gcttccaagg aactgcgcct 840
273 gttgatgcct tccctcccaa tggttatggc ttatacaaca tagtggggaa cgcattggga 900
275 tggacttcag actggtggac tgttcatcat tctgttgaag aaacgcttaa cccaaaaggt 960
277 ccccttctg ggaagaccg agtgaagaaa ggtggatcct acatgtgcca taggtcttat 1020
279 tgttacaggt atcgctgtgc tgctcggagc cagaacacac ctgatagctc tgcttcgaat 1080
281 ctgggattcc gctgtgcagc cgaccgcctg cccaccatgg ac 1122
284 <210> SEQ ID NO: 4
285 <211> LENGTH: 2130
286 <212> TYPE: DNA
287 <213> ORGANISM: Homo sapiens
289 <400> SEQUENCE: 4
290 acatggcccg cgggacaaca tggctgcgcc cgcactaggg ctggtgtgtg gacgttgccc 60
292 tgagctgggt ctcgctctct tgctgctgct gctctcgctg ctgtgtggag cggcagggag 120
294 ccaggaggcc gggaccggtg cgggcgcggg gtcccttgcg ggttcttgcg gctgcggcac 180
296 gccccagcgg cctggcgccc atggcagttc ggcagccgct caccgatact cgcgggaggc 240
298 taacgctccg ggccccgtac ccggagagcg gcaactcgcg cactcaaaga tgggtcccat 300
300 ccctgctgga gtatttaca tgggcacaga tgatcctcag ataaagcagg atggggaagc 360
302 acctgcgagg agagttacta ttgatgcctt ttacatggat gcctatgaag tcagtaatac 420
304 tgaatttgag aagtttgtga actcaactgg ctatttgaca gaggtgaga agtttggcga 480
306 ctccctttgtc ttgaaggca tgttgagtga gcaagtgaag accaatattc aacaggcagt 540
308 tgcagctgct ccctggtggt tacctgtgaa aggcgctaac tggagacacc cagaagggcc 600
310 tgactctact attctgcaca ggcgggatca tccagttctc catgtgtcct ggaatgatgc 660

```

RAW SEQUENCE LISTING

DATE: 03/14/2007

PATENT APPLICATION: US/10/775,678A

TIME: 10:46:31

Input Set : A:\10278-048001.txt

Output Set: N:\CRF4\03142007\J775678A.raw

```

312 ggttgccctac tgcacttggg caggggaagcg gctgcccacg gaagctgagt gggaatacag 720
314 ctgtcgagga ggccctgcata atagactttt cccctggggc aacaaactgc agcccaaagg 780
316 ccagcattat gccaacattt ggcagggcga ttttcgggtg accaactctg gtgaggatgg 840
318 cttccaagga actgcgcctg ttgatgcctt cctcccaat gggtatggct tataacaacat 900
320 agtggggaac gcatgggaat ggacttcaga ctgggtggact gttcatcatt ctgttgaaga 960
322 aacgcttaac ccaaaaggct ccccttctgg gaaagaccga gtgaagaaag gtggatccta 1020
324 catgtgccat aggtcttatt gttacaggta tcgctgtgct gctcggagcc agaacacacc 1080
326 tgatagctct gcttcgaatc tgggattccg ctgtgcagcc gaccgcctgc ccaccatgga 1140
328 ctgacaacca agggtagtct tccccagtc aaggagcagt cgtgtctgac ctacattggg 1200
330 ctttcctcag aactttgaac gatcccatgc aaagaattcc caccctgagg tgggttacat 1260
332 acctgcccac tggccaaagg aaccgccttg tgagaccaa ttgctgacct gggtcagtgc 1320
334 atgtgcttta tgggtgtggg catctttgga gatcatcacc atattttact tttgagagtc 1380
336 tttaaagagg aaggggagtg gagggaaacc tgagctaggc ttcaggaggc ccgcatccta 1440
338 cgcaggctct gccacagggg ttagacccca ggtccgacgc ttgaccttcc tgggcctcaa 1500
340 gtgccctccc ctatcaaag aaggaatgga cagcatgacc tctgggtgtc tctccaactc 1560
342 accagtctta aaaaggggat cagattctat tgtgacttca tagaatttat gatagattat 1620
344 tttttagcta ttttttccat gtgtgaacct tgagtatac taatcatgta aagtaagagt 1680
346 tctcttatgt attatgttcg gaagaggggt gtgggtgact ctttatattc gtactgcact 1740
348 ttgtttttcc aaggaaatca gtgtctttta cgttggtatg atgaatcca catggggccg 1800
350 gtgatgggat gctgaagttc agccgttgaa cacataggaa tgtctgtggg gtgactctac 1860
352 tgtgctttat cttttaacat taagtgcctt tgggttcagag gggcagtcac aagctctggt 1920
354 tccccctctc cccaaagcct tcagcgaacg tgaaatgtgc gctaaacggg gaaacctggt 1980
356 taattctaga tatagggaaa aaggaacgag gaccttgaat gagctatatt cagggtatcc 2040
358 ggtatattgt aataggggat aggaaacctt gttggctgtg gaatatccga tgctttgaat 2100
360 catgcaactgt gttgaataaa cgtatctgct 2130

```

363 <210> SEQ ID NO: 5

364 <211> LENGTH: 374

365 <212> TYPE: PRT

366 <213> ORGANISM: Homo sapiens

368 <400> SEQUENCE: 5

```

370 Met Ala Ala Pro Ala Leu Gly Leu Val Cys Gly Arg Cys Pro Glu Leu
371 1 5 10 15
374 Gly Leu Val Leu Leu Leu Leu Leu Ser Leu Leu Cys Gly Ala Ala
375 20 25 30
378 Gly Ser Gln Glu Ala Gly Thr Gly Ala Gly Ala Gly Ser Leu Ala Gly
379 35 40 45
382 Ser Cys Gly Cys Gly Thr Pro Gln Arg Pro Gly Ala His Gly Ser Ser
383 50 55 60
386 Ala Ala Ala His Arg Tyr Ser Arg Glu Ala Asn Ala Pro Gly Pro Val
387 65 70 75 80
390 Pro Gly Glu Arg Gln Leu Ala His Ser Lys Met Val Pro Ile Pro Ala
391 85 90 95
394 Gly Val Phe Thr Met Gly Thr Asp Asp Pro Gln Ile Lys Gln Asp Gly
395 100 105 110
398 Glu Ala Pro Ala Arg Arg Val Thr Ile Asp Ala Leu Tyr Met Asp Ala
399 115 120 125
402 Tyr Glu Val Ser Asn Thr Glu Phe Glu Lys Phe Val Asn Ser Thr Gly
403 130 135 140
406 Tyr Leu Thr Glu Ala Glu Lys Phe Gly Asp Ser Phe Val Phe Glu Gly

```

RAW SEQUENCE LISTING ERROR SUMMARY
PATENT APPLICATION: US/10/775,678A

DATE: 03/14/2007
TIME: 10:46:32

Input Set : A:\10278-048001.txt
Output Set: N:\CRF4\03142007\J775678A.raw

Please Note:

Use of n and/or Xaa have been detected in the Sequence Listing. Please review the Sequence Listing to ensure that a corresponding explanation is presented in the <220> to <223> fields of each sequence which presents at least one n or Xaa.

Seq#:32; Xaa Pos. 1,2,3
Seq#:79; Xaa Pos. 3,4,6
Seq#:81; N Pos. 590,626
Seq#:82; N Pos. 690,755
Seq#:83; N Pos. 6,47,81

VERIFICATION SUMMARY

DATE: 03/14/2007

PATENT APPLICATION: US/10/775,678A

TIME: 10:46:32

Input Set : A:\10278-048001.txt

Output Set: N:\CRF4\03142007\J775678A.raw

L:16 M:270 C: Current Application Number differs, Replaced Current Application Number
L:3704 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:32 after pos.:0
L:6087 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:79 after pos.:0
L:6155 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:81 after pos.:540
M:341 Repeated in SeqNo=81
L:6199 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:82 after pos.:660
M:341 Repeated in SeqNo=82
L:6226 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:83 after pos.:0
M:341 Repeated in SeqNo=83